

Teacher Evaluations

Cheyenne McCarthy, Tyler Oren, and John Buys

Introduction to Teacher Evaluation

A new teacher stands in front of the classroom. The awkward newness is beginning to wear off after four months. Dressed more casually for the frigid winter months of heatless hallways and colder classrooms, the teacher rubs their hands encouraging the students whose names the teacher finally knows to sit down and take out their homework from the night before.

Another unexceptional day in the paradise of public education is about to begin, when unannounced a suited administrator welcomes themselves into the classroom. Clipboard in hand and wrist-watch ticking the administrator finds a seat and waits expectantly for the novice to escape the frozen coma of shock at the newest arrival.

Once freed from the evil curse, the teacher begins the lesson doing their best to ignore the uninterested administrator who glances back and forth between wrist-watch and classroom clock until ten minutes has lapsed. Lazily the administrator checks boxes on his clipboard, gets up and moves on. Everyone breathes a sigh of relief. Another evaluation done, everything can go back to normal.

Unfortunately, this vignette is the all-too familiar model of teacher evaluation. A mandatory step in an unimaginative process, teacher evaluation had little meaning for many years. No Child Left Behind's narrowed definition of the highly qualified teacher as one, having a bachelor's degree, two, full state certification or licensure, and three, proven knowledge of subject material, created a significance for what it meant to be a professional in education (ed.gov).

A recent wave of teacher evaluation reform has spread across schools, districts and states inspired by the Bill and Melinda Gates Foundation's \$45 million "Measures of Effective Teaching Project" and the New Teacher Project's "Teacher Evaluation 2.0."

According to Robert Marzano (2012) teacher evaluation has two major focuses: measurement and development (p. 15). Measurement focuses quantify the teacher's abilities used within the classroom in a similar fashion to how a grade on a final works. Development focuses identify what a teacher performs best and how they need to alter their practice in order to become better educators. The best evaluations for teachers should do both; however, in practice many fall short, especially in terms of providing authentic opportunities for development.

The tug-of-war between measurement and development may seem like a simple fix, find an evaluation system that uses both and employ that, but as research suggests the battle is more persistent permeating into various hot-button issues. Part of the problem exists in how complex the practice of education is. With pressures from outside agencies like the Gates Foundation and the New Teacher Project, teacher retention and promotion are beginning to be attached to evaluation results that are determined by controversial value-added measures with merit pay rewards. With rival evaluation systems vying for relevance, accountability and inter-rater reliability comes into question. As this report will indicate, there are research-based systems and solutions out there. How to create a system of evaluation in lieu of competing opinions and research complicates the balance of an evaluation system

that can both measure and develop the professional.

Types of Evaluation

Observation

Bambrick-Santoyo (2012) identifies that part of the complexity of evaluations exists in defining the complexity of *what* goes on in the classroom where “Robert Marzano and his colleagues focus on 41 key strategies. Charlotte Danielson has suggested a framework for good teaching that encompasses a whopping 76 criteria” and *who* can adequately recognize these criteria in action (p. 28). Marzano’s 41 strategies and behaviors are broken down into three major categories: routine strategies, content strategies and strategies enacted on the spot (Marzano, p.17). Another popular model Rapid Assessment of Teacher Effectiveness (RATE) has only to behaviors including:

providing clear lesson objectives, understanding students background and comfort with material, using more than one delivery mechanism, providing multiple examples, providing appropriate nonexamples, maintaining an effective pace, providing students with feedback, engaging in timely use of guided practice, explaining important concepts clearly, keeping students actively engaged (Marzano, p. 16).

Part of the problem also exists within the amount of time given to observations. The MET project encourages four visits a year (three announced and one unannounced) and multiple observers with the same rubric (Marshall, p. 50). Other programs encourage the same observer to sample for ten-fifteen minutes ten unannounced meetings followed by an observer write up and follow up time between observer and teacher (Marshall, p. 51). All of these are improvements over the one drop-by so many principals had used in the past.

Traditionally, administrators (either the principal or the vice-principal) were in charge of evaluating teachers through observations. With changing evaluation systems are changing roles of observers including peer-observation, professional observers, and student input are all observation methods being encouraged in various systems.

Danielson (2012) identifies three key components to effective observation: observations based on evidence of “words spoken,” “actions,” and “appearance of the classroom,” not interpretations; “interpreting evidence against levels of performance” where the results are consistently quantifiable across various observers; and finally, “attention to the interaction skills of professional conversation, inviting teachers to reflect on practice and strengthen it in ways described by the instructional framework they use” (pp. 35-36).

Johnson and Fiarman (2012) explore Peer Assistance and Review (PAR) observers who go through a rigorous process in order to become consulting teachers who are hired to “leave the classroom for 3-5 years to provide intensive, individualized help to a caseload of 15-20 teachers” (p. 21) The results of this thoroughly developmental approach resulted in first-year teacher retention rates averaging around 90 percent and substantial savings for districts (Johnson & Fiarman, pp. 21-22).

Research into teacher evaluations with student observations have proven the overall effectiveness where Marshall (2012) observes that “students taught by teachers with student ratings

achievements achieve a full semester better than teachers whose students get low ratings” (p. 52). Also, the MET project has identified the “student perceptions of a given teacher’s strengths and weaknesses are consistent across the different groups of students they teach” giving a level of accountability to student input and observation for teacher evaluation (Gabriel and Allington, p.47) That being the case, Marshall is realistic in addressing concerns over high-stake student surveys which may not fully address appreciation for “tough, demanding teachers,” establishing that there are certain weaknesses in this evaluator approach as well where “although high student rating correlate with high achievement in the year surveyed, they correlate less well with success in the next grade or course “(p. 52).

Value-Added Measures

The United States’ education system’s zeal for testing is well recognized. This zeal has permeated the world of teacher evaluation. In what is probably the most controversial part of modern teaching evaluation, the inclusion of value-added measures has divided the educational world on teacher evaluation. Di Carlo (2012) defines value-added measure as a “specific type of growth model, a diverse group of statistical techniques to isolate a teacher’s impact on his or her students’ testing progress while controlling for other measurable factors, such as student and school characteristics that are outside that teachers control” (p. 38). Essentially, value-added measures include students’ standardized testing data in the teacher evaluation.

The percentages of value-added measures vary in size from as high as 50 percent in Tennessee’s Teacher Advancement Program with other systems using value added measures as low as 10-20 percent (Pieczura, p. 70). That being the case, standardized test data is rarely indicative of the best teaching; Goodwin and Miller (2012) state that “teachers account for only about 13 percent of the variance in student achievement,” 80 percent of the variance was found to be affected by student variables that include “home environment, student motivation, and prior knowledge” (p. 80). This issue becomes even murkier when addressing the publicizing of teachers’ value-added measures as both New York City and Houston teachers were subjected too. Pallas (2012) shared results that showed that the grading scales attributed with the value-added scores resulted in criminalizing teachers, who the prior year had been teachers of the year, with D and F grades (p. 56). Nor, did these measurements take into account affective factors such as English Language Learner influxes or how many students had been moved into the classroom saying absolutely nothing in terms of the prior learning that occurred at the other schools (Pallas, 56). Similar results were evident in the Maine Department of Educator’s grading system for Maine schools which only reiterated what everyone knew about the state’s schools: poorer schools with larger populations of students participating in Free and Reduced Lunch programs received lower grades than those with more affluent students and funding.

Proponents of value-added measures such as the Bill and Melinda Gates Foundation’s MET Project (2010) has identified that “in every grade and subject, a teacher’s past track record on value added-measures is among the strongest predictors of their students’ achievement gains in other classes and academic years” and that “teacher with high value-added scores on state tests tend to promote deeper conceptual learning as well” (p. 47). With both sides effectively finding research to debunk the

each other, the pressure of value-added measures will not be going anywhere.

Maine and Teacher Evaluation

Due to Federal requirements (and other factors), teacher evaluations have become more common in Maine over the past few years. Mainly being used for evaluation programs are the Board for Professional Teaching Standards (NBPTS), the Danielson Framework for Professional Teaching, and the Marzano Teacher Evaluation Model. All three of these programs encompass the standards for teacher evaluation posed by the Interstate Teacher and Support Consortium (InTASC). With these professional improvement programs in place, teachers seen as “ineffective” two years in a row could be dismissed. (Maine Education Association, 2009-2013)

In March 2013, Maine provisionally adopted Chapter 180, which consists of the rules for implementing required teacher and principal evaluation systems. This adoption, however, came after the hearing deadline which was in December 27. The Department of Education must now wait until the full Legislature approves this new set of rules so they can finally take effect. Superintendents throughout the state are working towards this implementation now – “the first deadline is to have a developed system during the upcoming school year (2013-14), followed by piloting the system in 2014-15, and fully implementing in 2015-16.” (Bowen, 2013)

With this push Education Evolving: Maine’s Plan for Putting Learners First was created. The second core priority stated in this document is titled Great Teachers and Leaders. Educators and administrators are working towards becoming highly effective teachers and school leaders by taking action in four main steps: 1) Common standards for teacher and leader effectiveness, 2) Initial preparation and ongoing professional development programs that are rigorous, relevant, and data-driven, 3) Next-generation evaluation systems for teachers and leaders, and 4) Communities of practice designed to foster continuous improvement.

First, a definition of what effective teaching and school leadership looks like was created. With the Maine Learning Results, there were standards for what students should be able to do and know, but not for the educators. Forming standards in which it shows what a teacher should be able to do would help establish teacher effectiveness. Maine’s goal is for a common understanding of effective teaching and leadership to shape the preparation, training, and evaluation of educators. The objective is to align a state statute and rules upon the adoption of the state standards for teacher and leader effectiveness.

Because of this demand for the meeting of the teacher and leader effectiveness standards, there has been a push in ongoing professional development programs. These programs will provide leadership training and development, creating more effective educators and leaders. Maine’s goal is to have educators being consistently supported through high-quality training and professional development. The objective is to expand the access of teachers and school leaders to initial and ongoing high-quality training and professional development.

Based on The Widget Effect, the 2009 study of educator evaluation systems, The New Teacher Project came to the conclusion that the educator evaluation systems being used “fail to differentiate performance among teachers.” This is because teacher effectiveness, which is a key aspect in improving student achievement, was not measured, recorded, or even used to promote decision-making in any

meaningful way. This prompted the state to work with districts to create next-generation evaluation systems and “develop regional teacher development centers that not only support the training of the evaluators themselves, but make use of evaluation data to design and implement targeted professional development.” (Department of Education, 2013) Maine’s goal was/is to have highly effective educator evaluation systems in place within every school district in the state. The objective is to “adopt statewide guidelines for locally developed teacher and leader evaluation systems, and support the development of a network of trained evaluators based in regional teacher development centers.” (Department of Education, 2013)

Lastly, a collaborative platform is being developed so that visitors can “browse the work of various practice groups, participate in conversations about the materials and educational practice challenges, and join practice groups where they can more actively participate in ongoing development of education solutions.” (DOE, 2013) This platform could become the home of effective instructional practices and other useful resources to educators and school leaders. Maine’s goal is to have educators participate easily and often in statewide sharing of instructional best practices and professional development opportunities. The objective is to develop a state-level, online resource center that will be solely for effective educational practices and professional development resources to be shared.

Self Assessment and Charlotte Danielson

Self-awareness and self-assessment are valuable tools within the teacher’s kit. These techniques guide teachers along as they develop their own potential as educators, in turn making their students more prepared learners. Furthermore a teacher who is experienced with self-reflection and self-assessment can pass these skills along to their students, helping them to master their own educational career more fully. As an instructor reflects on their lesson plan and material within they can begin by reflection on the standards that their content will allow their students to meet, and how their course goals complement those standards. Ideally these reflections should occur before actual instruction to ensure a smooth delivery on students. Following instruction educators can move to reflection of how their students reacted to the material and whether or not they were actually able to meet their expected goals and standards.

There is an important correlation between perceived efficacy and actual efficacy that has been found in several studies (Goddard, Hoy, & Hoy, 2004; Herman, Meece, & McCombs, 2000; Mascall, 2003; Muijs & Reynolds, 2001; Ross, 1992; Ross & Cousins, 1993), meaning that the more effective a teacher considers their instruction actually is the more effective that instruction is on their students. Furthermore the studies found that the most successful tactic in raising perceived efficacy was to practice self-evaluation.

An important aspect of developing accurate self-assessment comes from participating in peer evaluation, both through observing your own peers and welcoming observations from them. Through their observations peers can help focus and refine the self-reflective process. This practice is made even more effective if colleagues are able to evaluate student goals and outcomes. Three sources of feedback from colleagues have been highlighted by Bandura (1997) which include, social persuasion, convincing colleagues that they are able to rise to a challenge, vicarious experience, which describes highlighting the

successes of those similar to the instructor, and finally managing physiological and emotional states, strengthening positive feelings arising from teaching and interpreting them as indicative of teaching ability or reducing negative feelings arising from teaching, such as stress.

Ideally teacher evaluation should take on a holistic approach to generate the most accurate feedback for the teacher. Evaluation should consist of not only the traditional observation conducted by a school administrator or department head, but by numerous colleagues and faculty members, the instructor's students, parents and self-reflection. The reasoning here is similar to the movement to transition away from high stakes testing , being that a single test can be skewed by a number of variables that can effect performance. This same phenomenon can manifest itself in instructors as well. Charlotte Danielson writes in her book, *Enhancing Professional Practice*, a rather comprehensive framework for teaching and self-evaluation. Danielson presents the primary reason for teacher evaluation as a quality assurance placed on educators, however her reason given to continue teacher evaluation is to change the paradigm associated with evaluations from a quality assurance to a means to promote professional development in educators.

Overall our findings blend well with Charlotte Danielson's research. Primarily good teacher evaluation needs to possess several key factors. Evaluation must contain a clear definition concerning what "good" teaching actually is. Evaluation must be able to assess evidence and artifacts as proof of effective teaching. One of the most important facets of effective evaluation is that evaluators must be trained to give similar results, for example two evaluators should not give widely varied evaluations of the same instructor giving the same lesson. Instructors must also be presented with professional development to make them familiar with the evaluation process. Finally Danielson and our group agree that for good evaluation to occur there must be a concrete process for final judgment of evaluation to begin.

References

- Bambrick-Santoyo, Paul (2012, November). Beyond the Scoreboard. *Educational Leadership*, 70 (3), 27-30.
- Bowen, Stephen. (2013, May). Update on status of teacher/principal evaluation rules. Maine Department of Education. Retrieved November 22, 2013, from <http://mainedoenews.net/2013/05/10/update-on-status-teacherprincipal-evaluation-rules/>
- Danielson, Charlotte. (2012, November). Observing Classroom Practice. *Educational Leadership*, 70 (3), 32-37.
- Danielson, Charlotte. *Enhancing Professional Practice*. 2nd ed. Alexandria, VA: ASCD, 2007. 96-102. Print.
- Di Carlo, Matthew. (2012, November). How to Use Value-Added Measures Right. *Educational Leadership*, 70 (3), 38-42.
- Hambleton, Paul. (2009-2013). Teacher Evaluations: What You Need to Know. maine.nea.org.

- Retrieved November 22, 2013, from <http://maine.nea.org/home/1356.htm>
- Gabriel, Rachael, & Allington, Richard. (2012, November). The MET Project: The Wrong \$45 Million Question. *Educational Leadership*, 70 (3), 44-49.
- Goodwin, Bryan & Miller, Kirsteb. (2012, November). Use Caution with Value-Added Measures. *Educational Leadership*, 70 (3), 14-19.
- Johnson, Susan Moore, & Fiarman, Sarah E. (2012, November). The Potential of Peer Review. *Educational Leadership*, 70 (3), 20-25.
- Maine Department of Education. (2013). Education Evolving: Maine's Plan for Putting Learners First Core Priority Area 2: Great Teachers and Leaders. Retrieved November 22, 2013, from http://www.maine.gov/doe/plan/education_evolving/cpa2.html
- Marshall, Kim. (2012, November). Fine-Tuning Teacher Evaluation. *Educational Leadership*, 70 (3), 50-53.
- Marzano, Robert J. (2012, November). The Two Purposes of Teacher Evaluation. *Educational Leadership*, 70 (3), 14-19.
- Pallas, Aaron M. (2012, November). The Fuzzy Scarlet Letter. *Educational Leadership*, 70 (3), 54-57.
- Pieczura, Michelle (2012, November). Weighing the Pros and Cons of TAP. *Educational Leadership*, 70 (3), 70-72.